Exceptional Children
2022, Vol. 88(2) 205–222

DOI: 10.1177/00144029211062588
journals.sagepub.com/home/ecx
SAGE

First Step Next and homeBase: A Comparative Efficacy Study of Children With Disruptive Behavior

Andy J. Frey[1], Jason W. Small[2], John R. Seeley[3,5], Hill M. Walker[3,5], Edward G. Feil[2], Jon Lee[5], Dana Cohen Lissman[3], Shantel Crosby[1], and Steven R. Forness[4]

Abstract

Disruptive behavior disorders in childhood are increasingly pervasive and associated with numerous, negative long-term outcomes. The current study examined whether adding a brief, home-visitation intervention to an existing, multi-component (child and teacher) intervention, would improve social-emotional and behavioral outcomes for young children with challenging behavior in home and school settings who required intensive support. A total of 379 teacher-parent-student triads were screened for elevated levels of behavioral risk in school and home settings and then randomly assigned to school only intervention (i.e., teacher and student components), home only intervention (i.e., parent), both combined, or business-as-usual control conditions. We examined baseline and posttest outcomes across prosocial behavior, problem behavior, and academic domains. The results demonstrated substantial support for the teacher and child-focused condition and combined conditions, and modest support for the parent-focused condition. The study advances the literature by increasing the knowledge base related to these interventions implemented alone and in combination.

Keywords

behavior disorders, early intervention, comparative efficacy, home visitation, parenting

Disruptive behavior in childhood is characterized by symptoms such as temper tantrums, oppositional behavior, and interpersonal aggression that not only result in persistent impairment but are also often associated with comorbid externalizing disorders such as attention deficit hyperactivity disorder (ADHD) and conduct disorder (CD; Bradshaw et al., 2010). Prevalence rates of disruptive behavior are continually increasing, with 5.3% of youngsters receiving at least one out-of-school suspension (Noltemeyer et al., 2015; U.S. Department of Education, 2018). In the past decade prevalence of clinically significant mental health concerns for young children have been relatively steady, ranging between 8%–12% (Blanchard et al., 2006; Forness et al., 2012; Zaim & Harrison, 2020), with disruptive behavior disorders

[1]University of Louisville Kent School of Social Work
[2]Oregon Research Institute
[3]University of Oregon, College of Education
[4]University of California
[5]University of Northern Arizona College of Education

Corresponding Author:
Andy J. Frey, Kent School of Social Work, University of Louisville, Louisville, KY 40292.
Email: afrey@louisville.edu

being among the most frequent and having the most significant long-term effect. Thus, implementing early interventions is of critical importance in addressing these disorders successfully.

Multicomponent Interventions for Disruptive Behaviors

There are dozens of randomized controlled trials (RCT) of psychosocial interventions to address this issue. However, a review of 28 of these studies found only two that involved teachers as the primary program implementer (Kaminski & Claussen, 2017). Interestingly, even in meta-analyses of RCTs that primarily involve parent interventions, only about one-fourth (Granski et al., 2020) to one-third (Dretzke et al., 2009) directly involve children themselves as targets of the intervention sessions. Mean effect sizes were rather impressive, in the range of 0.7 to 0.9; but the sample sizes of most of these RCTs were relatively small, with group size averaging 21 participants. Additionally, Epstein et al. (2015) found that parent only and multi-component interventions with a parent component both had mean effect sizes of 1.2. Thus, RCTs for the vast majority of these psychosocial interventions appear to involve parent training or consultation, with relatively few devoted specifically to the classroom with the teacher as the interim interventionist and the student behavior as the primary target of the intervention.

Comparative Efficacy Studies

More research is needed comparing the relative merits of different mechanisms of multi-component interventions, or combinations of interventions addressing the child, teacher/classroom, and parent subject groups. It is particularly important to examine combinations of interventions that are deployed by school personnel because 1) most children who receive mental health services receive them in schools (Seeley et al., 2014; Vidair et al., 2014); and 2) their estimated costs, implementation effort, and efficacy can drive their usage. Comparative efficacy or effectiveness studies in which one intervention is directly compared with another or with a combination of interventions remain exceedingly rare in special education and school mental health research (Forness et al., 2014; Walker, Forness, & Lane, 2014; Walker, Severson, & Feil, 2014). In one of the few examples, Webster-Stratton et al. (2004) found, in a study of their Incredible Years program, that parent-only training had a larger influence on child negative behavior at home (approximately 0.7 effect size) than a combination of parent and teacher training (effect size of 0.4); unfortunately, there was no direct comparison of parent training only versus teacher training only components. In another example, a relatively large study (N = 1,500) of children compared additive effects of classroom-wide, small group, and parent training components (Group, 2002). Results showed that only children receiving all three components improved in disruptive behavior significantly compared to the control group. However, a review of 47 RCTs and quasi-experimental comparisons of school mental health interventions included only 12 with both teacher and parent components (Vidair et al., 2014), and none included direct comparisons among components.

Current Study

The current study sought to establish whether adding a brief, home-visitation intervention to an existing, multi-component (child and teacher components) intervention, would improve social-emotional and behavioral outcomes for young children with challenging behavior in home and school settings who required intensive supports. The multi-component intervention, First Step to Success, is a fully manualized, Tier 2 intervention designed for students in preschool through second grade with disruptive behaviors (Walker et al., 2015). This intervention has been the focus of both small- and large-scale randomized controlled trials (Feil et al., 2014; Feil et al., 2021; Sumi et al., 2013; Walker et al., 1998; Walker et al., 2009) and is listed in the What Works

Clearinghouse (WWC) as having positive effects on externalizing behavior problems and potentially positive effects on emotional/internalizing behavior, social outcomes, and academic performance.

The original First Step to Success intervention (Walker et al., 1997) contained a home visitation component focusing on positive parenting skills, whereas the recently revised version, First Step Next (FSN) replaced the home visitation component with home-school communication strategies to engage parents (Walker et al., 2018). Given the recent FSN revision that removed the home visitation component, an examination of FSN outcomes—individually and in combination with a brief home visitation intervention—was warranted. The parent-focused intervention, homeBase (hB), seeks to improve the social, behavioral, and academic outcomes of young children by providing home-based support to caregivers. Small-scale, quasi-experimental studies have shown the hB intervention can be implemented with fidelity, is a socially valid intervention (Frey et al., 2019), and has promising effects when delivered with the original First Step intervention (Frey et al., 2015).

The purpose of the present study, using a comparative efficacy approach, was to examine the effects of FSN alone, hB alone, and FSN + hB, compared to one another and to a business-as-usual control condition. The primary goal of this study design was to inform decision making on how best to distribute resources for students with disruptive behavior disorders who require intensive supports.

Method

Participants

There were 379 teacher-parent-student triads from 100 schools in five districts in Kentucky and Indiana that participated in the project. Participating teachers were predominantly female (93%) and White (85%). A small percentage of teachers (12%) reported their race as Black or African American. On average, teachers had worked in the field for 11 years ($SD = 8.29$) and held either a master's degree (71%) or bachelor's degree (28%). Participating parents had a mean age of 35 years ($SD = 9.33$ years) and were predominantly female (90%), and mostly Black (52%) or White (43%). Fourteen percent of parents held a bachelor's degree or higher. Nearly 70% of parents were currently working and just over one-third (36%) were living below the poverty level. Of the participating students, 71% were eligible for reduced-price meals at school, and students who met inclusion criteria were primarily male (74%) and either Black (52%) or White (37%).

Recruitment and Screening

After obtaining IRB approval for the study from the University of Louisville and all participating school district institutional review boards, project staff recruited 842 teachers across five cohorts from 2015–2020. We used teacher- and parent-reported screening data within a multi-step process to identify students eligible for inclusion in the project. At step one, teachers completed Stages 1 and 2, involving teacher rankings and ratings, from the Systematic Screening for Behavior Disorders (SSBD; Walker, Forness, & Lane, 2014; Walker, Severson, & Feil, 2014) to identify five students from their classrooms who were at elevated risk for externalizing behavior. Project staff rank ordered students who met SSBD Stage 2 cutoff criteria in terms of severity, and targeted the highest-ranked student in each classroom for inclusion. At step two, project staff collected the externalizing scale ratings of the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001) from the parents of the highest ranked student to verify the student was indeed struggling with behavioral issues across school *and* home settings. Project staff recruited a family to participate in the project if the student was in the borderline or clinical range on the CBCL externalizing scale. If the highest ranked student on the SSBD did not meet CBCL criteria or if we were unable to screen or recruit the highest ranked student, we repeated the process with

the next highest rank student in the classroom. The majority of participating students (87%) were in the clinical range on the CBCL externalizing scale and the rest were in the borderline range. Teachers and parents received $50 to complete the screening process. The CONSORT diagram in Figure 1 provides a breakdown of participants at each stage from consent to allocation and through collection of post-test data.

First Step Next and homeBase Interventions

FSN draws on support from multiple stakeholders (e.g., parents, teachers, peers), and includes three major tasks: direct social skills instruction; the green card game; and home-school connections (Walker et al., 2015). During social skills instruction, a behavioral coach helps the focus student forge positive personal relationships, develop problem-solving skills, and improve self-regulation through delivery of school skills lessons (i.e., Super Student Skills). During the green card game, a color-coded card functions as a tool for the teacher to provide subtle but direct and immediate, non-verbal feedback that encourages continued display of the Super Student Skills or encourages the child to stop, think, and change their behavior from inappropriate to appropriate. The green card game, which over the course of FSN implementation can be extended to cover the entire school day, is easily integrated into daily academic and social skills lessons as well as into other classroom activities. The focus child's peers also participate during the game and in the reinforcement activity if earned. For the home-school connection component, caregivers receive (a) daily feedback in the form of a note or phone call from the FSN coach, and (b) materials focused on promoting positive parenting strategies.

FSN delivery occurs across three phases. During the coach phase (program Days 1–7) a trained coach works directly with the focus student in delivering the social skills lessons, implementing the green card game, and communicating daily with the student's parents or caregiver via notes and phone calls. During the transition phase (program Days 8–10), the coach transitions control and management of the intervention to the teacher. In the teacher phase (program Days 11–30), the teacher facilitates playing of the green card game by determining when it will be played, managing the color-coded card in response to desirable or undesirable student behavior, and administering the group reward if the focus student meets criteria. Also, during the teacher phase, the teacher reviews the Super Student Skills individually with the focus child as needed and sends and receives the daily home note. During the teacher phase, the coach assists with troubleshooting if the student does not meet criteria and/or the teacher needs assistance.

The hB intervention consists of 3–6, 60-min home visits over several months. During the sessions, parents are encouraged to align their parenting practices consistent with one or more of the five universal principles of positive behavior support (Sprague & Golly, 2013): (a) establishing clear expectations; (b) directly teaching expectations; (c) reinforcing the display of expectations; (d) minimizing attention for minor inappropriate behaviors; and (e) establishing clear consequences for unacceptable behavior. Sessions are delivered within a multi-step process for increasing intrinsic motivation to adopt and implement an evidence-based practice with integrity. The four steps are:

1. Engage in values discovery.
2. Assess current practices.
3. Share performance feedback.
4. Offer extended consultation, education, and support.

During each step, the coach uses Motivational Interviewing (MI) to guide and strengthen the caregiver's engagement with and commitment to behavior change (Frey et al., 2015). MI is defined as "a collaborative conversational style for strengthening a person's own motivation and commitment to change" and has demonstrated efficacy when used alone and

in combination with other interventions (Miller & Rollnick, 2012, p. 12).

First Step Next and homeBase Coaches

There were 29 FSN coaches who were university employees and participated in FSN implementation. They ranged in age from 24 to 62 years old ($M[SD] = 43.50$ [10.30]) and were predominantly female (90%). Coaches reported their race as either Black (28%) or White (72%). More than half of the coaches held a master's degree or higher (63%). Project research staff trained all coaches using role-play activities and practice introducing and implementing FSN to all students and teachers. FSN behavioral coaches were trained via a 2-day workshop where they were introduced to the intervention material, viewed and reflected on videos of the intervention being implemented correctly, and role-played several of the procedures (e.g., the initial meeting with parents, delivering the social skills lessons to the target child, introducing the program to the target child, and implementing the 1st day of the program). Coach trainees then practiced implementing the intervention with at least one child who was not enrolled in the study under the supervision of an experienced implementer. Coaches also participated in weekly supervision meetings led by an experienced implementer with other coaches to troubleshoot during the study.

The hB coaches who participated in this study were also university employees. In total, 25 coaches participated in hB implementation. They ranged in age from 23 to 61 years old ($M[SD] = 35.54$[13.80]) and were predominantly female (83%). Coaches reported their race as either Black (21%) or White (79%). More than half of the coaches held a master's degree or higher (67%) and an additional 21% reported having completed at least one year of master's-level coursework.

Research staff, who were members of the Motivational Interviewing Network of Trainers, provided training to all participating hB coaches using the Motivational Interviewing Training and Assessment System (MITAS; Frey et al., 2017). The MITAS workshops were used to introduce school personnel to the core elements of MI and detailed how to integrate critical MI knowledge and skills into school-based work with teachers, students, and parents. The training involved 12 hr of didactic workshops and a simulated practice session, where the trainees implemented the hB intervention with an experienced implementor who portrayed a parent and provided individualized feedback to the coach. The simulated practice sessions were audio-recorded and coded using a measure of MI fidelity; however, there were no other formal criteria required prior to being assigned a family to work with. Coaches also participated in weekly supervision meetings led by an experienced implementor with other coaches to troubleshoot any problems during the study.

Data Collection Procedures

Prior to randomization, parents and teachers completed a baseline questionnaire containing demographic and outcome measures. Trained observers, who were blind to intervention conditions, collected in vivo observation data in participating classrooms. At post-test, parents and teachers completed a survey containing outcome measures, and those randomized to one of the intervention conditions completed another survey addressing program satisfaction, therapeutic alliance, and barriers to participation. Teachers and parents received US $75 for completing a questionnaire at each time point. A brief description of the measures collected; sample-specific data on internal consistency (e.g., coefficient alpha); and inter-rater reliability (IRR) information for observation measures is provided subsequently. We assessed IRR via two-way mixed effects, absolute agreement, average-measures intra-class correlations. This approach to examining IRR corrects for chance agreement and, therefore, does not overestimate agreement levels between observers (Hallgren, 2012).

Outcome Measures

Social Skills Improvement System Rating Scales (SSiS). The SSiS (Gresham & Elliott, 2008) was collected to assess the child's

social skills, problem behavior, and academic competence. The teacher-completed version consisted of 46 social skills items (α = .90), 30 problem behavior items (α = .74), and seven academic competence items (α = .96). The parent-completed version consisted of 46 social skills items (α = .93) and 33 problem behavior items (α = .82). Teachers and parents rated social skills and problem behavior items on a 4-point frequency scale ranging from "never" to "almost always" and academic competence items on a 5-point scale from "lowest 10%" to "middle 40%" to "highest 10%." We converted raw scores to standard scores using gender-specific normative data.

Systematic Screening for Behavior Disorders. We collected three Stage 2 Systematic Screening for Behavior Disorders (SSBD) measures (Walker, Forness, & Lane, 2014; Walker, Severson, & Feil, 2014): The Critical Events Index (CEI), the Adaptive Behavior Index (ABI), and the Maladaptive Behavior Index (MBI). The CEI (α = .81) used in this project included 30 dichotomous (i.e., yes/no) items assessing low-frequency behaviors associated with long-term adjustment difficulties. The 12-item ABI scale (α = .90) assessed a student's adaptive behavior and the 11-item MBI scale (α = .88) assessed a student's maladaptive behavior. Teachers rated the ABI and MBI items on a 5-point rating scale ranging from "never" to "frequently." We computed raw total scores for each measure. Students exceeded SSBD risk criteria if their CEI score was five or higher or if they had an ABI score of 30 or less and an MBI score of 35 or more.

Child Behavior Checklist and Teacher's Report Form. Parents completed five of six DSM-oriented scales from the CBCL and teachers completed three of five DSM-oriented scales from the Teacher Report Form (Achenbach & Rescorla, 2001). Both sets of informants reported on attention deficit/hyperactivity problems (ADHD), oppositional defiant problems (ODD), and conduct problems (CD). The teacher- and parent-reported ADHD scales had 13 items (α = .88) and seven items (α = .80), respectively; the CD scales had 13 items (α = .84) and 17 items (α = .82), respectively; and the ODD scales both had five items (α = .81 and α = .60, respectively). Coefficient alpha for the 13-item affective scale was .76 and, for the 6-item anxiety scale, it was .73. All items were scored on a 3-point scale. We examined the data based on borderline (1.5 *SD* above mean) and clinical (2 *SD* above mean) cutoffs specified by the test developers.

Student–Teacher Relationship Scale (STRS). We collected the 12-item conflict subscale (α = .87) from the STRS (Pianta, 2001). The conflict subscale assessed the conflict dimension of the teacher–student relationship, the perceived unpredictability of the student, and emotional exhaustion associated with the relationship. Teachers rated the items on a 5-point scale ranging from "definitely does not apply" to "definitely applies." Higher total scores indicated higher levels of conflict.

Academic Engaged Time. Project staff collected direct observation data on three separate days at baseline and post-intervention using the SSBD stage 3 measure of student Academic Engaged Time (AET; Walker, Forness, & Lane, 2014; Walker, Severson, & Feil, 2014). AET was collected as an estimate of the amount of time a student spends engaged in academic activities and provided an important indicator of academic success and adjustment to classroom expectations. Each observation was 20 min in length. We examined the mean percent of AET across the three observations. We collected IRR data on 12% of over 2,100 AET observations conducted over the course of the project. Based on Cicchetti and Sparrow's (1981) benchmarks for categorizing the quality of the ICC, IRR for the AET was excellent, ICC (3,1) = .96.

Process Measures

First Step Next Fidelity Checklist. We used an abbreviated version of the implementation fidelity checklist (Walker et al., 2009) to assess adherence and quality during coach

and teacher implementation of FSN. The 12-item fidelity checklist assessed tasks pertaining to general FSN implementation (three items), use of the green-and-red card (four items), delivery of points and feedback (three items), peer involvement (one item), and school–home connections (one item). The checklist assessed delivery (e.g., adherence using a dichotomous yes-or-no scale) and quality of delivery using a 5-point scale (i.e., 0 = "very poor," 0.25 = "poor," 0.50 = "okay," 0.75 = "good," and 1.0 = "excellent"; $\alpha = .87$). Observers conducted three fidelity checks during FSN implementation: once during the coach phase and twice during the teacher phase. IRR, collected on 20% of the FSN fidelity checks conducted, was excellent for FSN adherence (ICC [3,1] = .91), and less so for quality (ICC [3,1] = .77) scores.

homebase Fidelity. We used the Motivational Interviewing Treatment Integrity Tool (MITI 4.2; Moyers et al., 2016) to assess the coaches' MI quality during the hB sessions. The MITI coding system was used to examine the verbal behavior of the hB coaches delivering MI. We used the MITI to code 10 behavior categories including simple reflections (SR), complex reflections (CR), affirmations, and questions; and four global dimensions: cultivating change talk (CCT), softening sustain talk (SST), partnership, and empathy. Four summary scores were generated using the raw behavior counts and global scores: (a) relational skills; (b) technical skills; (c) the percent of CRs; and (d) the ratio of reflections to questions. The relational skills global summary score was calculated by computing the mean of the partnership and empathy items. The technical skills global summary score was calculated by computing the mean of CCT and SST. Percent of CRs was calculated by dividing the number of CRs by total reflections (i.e., SR + CR). Finally, the ratio of reflections to questions was calculated by dividing total reflections by the number of questions posed during a session. Moyers et al. (2016) have identified "Basic Competency" and "Proficiency" thresholds for each summary score.

We conducted fidelity checks of audio recorded hB sessions. An independent team of trained coders randomly selected a continuous 20-min sample from each tape (See Small et al., 2020 for additional procedural details). IRR checks were conducted on a random sample of 20% of sessions. IRR scores ranged from .61 to .80 for the global scores, behavior counts, and summary scores. In general, ICCs were comparable to those reported by the MITI developers (Moyers et al., 2016).

Program Dosage and Participant Compliance. For participants in the FSN intervention, we used coach- and teacher-completed classroom monitoring forms (CMFs) to calculate measures of program exposure (dosage) and student compliance. The coach and teacher (during their respective FSN implementation phases) used the CMF to record the focus student's daily participation in the program. At the end of implementation, CMFs provided documentation of the total number of program days completed, the points and rewards earned daily by the child, and the number of daily parent connection notes sent and returned. In accordance with previous studies of First Step (Sumi et al., 2013; Walker et al., 2009), we calculated FSN dosage as the proportion of program days, delivered out of the possible 30 available days, and student compliance as the proportion of program days successfully completed out of the total number of program days administered.

For the hB intervention, we used data from the hB Coach Checklist (Frey et al., 2015) to measure program dosage. During implementation of the hB intervention, the coach used the checklist to track the number of home visits, the number of steps completed, and the activities completed and tools used at each step. The number of home visits and steps completed were highly correlated ($r = .83$).

Program Effectiveness and Participant Engagement. FSN and hB coaches completed two measures to assess the perceived effectiveness of each program and participant engagement. FSN coaches completed a 3-item effectiveness scale ($\alpha = .88$) and a 9-item measure ($\alpha = .93$) assessing the

extent to which the teacher changed his or her behavior as a result of participation in the FSN intervention. Coaches who delivered the hB intervention completed a 6-item effectiveness scale ($\alpha = .91$) and an 11-item engagement scale ($\alpha = .94$). Coach-reported effectiveness and engagement items were rated on a 5-point Likert scale ranging from "strongly disagree" to "strongly agree."

Alliance. At post intervention, we collected alliance data from the coach, teacher, and parent to assess their partnership as it related to program implementation. The Therapeutic Alliance Scale has been used in previous large-scale efficacy trials (Sumi et al., 2013; Walker et al., 2009). Parents participating in the hB intervention completed an 18-item parent-coach alliance scale ($\alpha = .96$); teachers participating in the FSN intervention completed an 8-item teacher-coach alliance scale ($\alpha = .96$); coaches delivering the FSN intervention completed an 8-item teacher-coach alliance scale ($\alpha = .96$); and coaches delivering the hB intervention completed an 8-item parent-coach alliance scale ($\alpha = .97$). Respondents rated alliance on a 5-point frequency scale ranging from "never" to "always." Alliance items assessed approachability, communication skills, follow through, shared goals, willingness to collaborate, and overall effectiveness.

Social Validity. We collected data on satisfaction from parents randomized to the three intervention conditions. We also collected satisfaction data from teachers randomized to the FSN only or FSN + hB intervention conditions. Parents completed a 12-item satisfaction measure ($\alpha = .93$). Teachers completed a 13-item satisfaction measure ($\alpha = .91$). Satisfaction items for both sets of informants were scored on a 5-point Likert scale from "strongly disagree" to "strongly agree." Satisfaction items assessed usability, support, and program effectiveness.

Analysis

For each continuous outcome, we estimated a linear regression model using Mplus 7.0 statistical software (Muthèn & Muthèn, 1998–2010). Specifically, each outcome was regressed on three dichotomous dummy variables—one for each of the three intervention conditions—and one covariate (i.e., the baseline value of the outcome). We centered the baseline value of the outcome so that the intercept represents the mean score of the control condition on each outcome, and the effect for each of the intervention conditions represents the posttest difference between the control and intervention groups. For categorical outcomes (i.e., DSM-oriented scales), we estimated logistic regression models in Mplus 7.0. For each diagnostic scale, we calculated a dichotomous variable capturing improvement from baseline to posttest (e.g., movement from the clinical range to borderline or normative range, or movement from the borderline range to normative range). Thus, the reported odds ratio for each categorical outcome indicated the probability of participants randomized to intervention conditions demonstrating improvement on diagnostic outcomes at post intervention.

We used multiple imputation techniques to account for missing data. We generated 40 imputed datasets in Mplus using an imputation model that included baseline and posttest scores on all outcomes of interest, and six auxiliary variables as potential correlates of missingness. Given that rates of missing data were higher for participating parents, we included auxiliary variables in the models that have been shown to be predictive of subsequent dropout, and which are associated with higher levels of familial stress (Reinke et al., 2012). The six auxiliary variables included in the imputation model were: (a) student rank on the SSBD; (b) student's biological sex; (c) participating parent's marital status; (d) parent's education level; (e) estimated annual household income; and (f) number of children in the household.

In accordance with the WWC (U.S. Department of Education) recommendations, we applied the Benjamini–Hochberg (B-H) correction to adjust for multiple within-domain comparisons (Benjamini & Hochberg, 1995). To calculate a B-H correction, outcomes were ranked in ascending order within domain based on p values and a

cutoff for each is calculated. For the pro-social behavior domain, rank-ordered intervention effects for the three outcomes were considered significant at a .05 alpha level if p values were less than .017, .033, and .050, respectively. For the problem behavior domain, rank-ordered intervention effects for the four outcomes were significant at a .050 alpha level if p values were less than .013, .025, .038, and .050. Finally, for the two outcomes in the academic domain, rank-ordered intervention effects were significant at the .050 level if p values are less than .025 and .050. As a measure of effect size, we reported Hedges' g. Small, medium, and large effect sizes correspond to a Hedges' g of .20, .50, and .80, respectively. As a measure of practical significance, we also reported the WWC improvement index, which represents the expected change in percentile rank for an average student in the control group if the child had received the intervention.

Results

Baseline Equivalence

We examined the equivalence of student, teacher, and parent demographics and student outcomes at baseline. As reported in Table 1, there were no statistically significant differences between students in each of the conditions on a range of student demographic and screening variables. Additionally, there were no statistically significant differences at baseline between conditions for the three pro-social outcomes, four problem behavior outcomes, and two academic outcomes. We examined equivalence for the following baseline parent demographic variables: age, percent female, percent Black, percent White, percent with a bachelor's degree or higher, percent currently employed, and percent below poverty level.

There were no statistically significant differences between participating parents on these demographic characteristics. At baseline, participating teachers did not differ on the number of years teaching, number of years at their current school, education level, percent female, or percent Black. They did differ on the percent who reported their race as White. As compared to the FSN only (90%), hB only (88%), and FSN + hB (88%) conditions, significantly fewer teachers randomized to the control conditions (75%) reported their race as White ($\chi^2[3] = 11.62$, $p = .009$).

Attrition and Missing Data

We obtained at least one type of post data (i.e., observation, parent-report, or teacher-report) from 95% of participants. We collected post-intervention AET observation data for 86% of students and collected post-intervention survey data from 92% of teachers and 81% of parents. We used a two-step approach to test the assumption that data were missing completely at random (MCAR). We first examined patterns of missing data and Little's MCAR test, a global test of MCAR. In turn, we also conducted univariate t tests for continuous variables and contingency table analysis for categorical variables to examine, for each outcome, whether cases with missing data differed from those without missing data. We conducted these additional analyses on relevant variables including child and parent demographics, and baseline values on screening and outcome measures. Little's MCAR test was nonsignificant ($\chi 2 = 851.06$, $n = 379$, $p = .137$) and none of the examined variables were significantly associated with missing data groups, suggesting that data were MCAR.

Process Data

FSN Fidelity, Dosage, and Compliance Data. Adherence to FSN's core components was high during the coach phase (M [SD] = .99[.04]) and teacher phase (M [SD] = .98 [.07]). Quality of implementation was also high during both phases. Mean quality ratings during the coach phase were .96 (SD = .07), and .90 (SD = .09) during the teacher phase. The mean percentage of programs days (out of 30) completed by students was 66% (SD = .27). Eight percent of participants did not complete any program days and 11% completed all program days. Overall,

Table 1. Baseline Equivalence of Student Demographic and Screening Characteristics.

	Total (n = 379)	Control (n = 95)	FSN only (n = 94)	HB only (n = 96)	FSN + HB (n = 94)	Test statistic	p-value
Demographic Characteristic							
Age *M* (*SD*)	6.75 (1.23)	6.80 (1.27)	6.87 (1.27)	6.58 (1.15)	6.77 (1.22)	0.96	.414
Percent female	98 (25.86)	23 (24.21)	26 (27.66)	22 (22.92)	27 (28.72)	1.13	.770
Percent African American	195 (51.45)	47 (49.47)	50 (53.19)	45 (46.87)	53 (56.38)	1.98	.576
Percent Caucasian	140 (36.94)	39 (41.05)	36 (38.30)	34 (35.42)	31 (32.98)	1.49	.684
Grade level						8.28	.507
Kindergarten	94 (24.80)	21 (22.11)	21 (22.34)	26 (27.08)	26 (27.66)		
1st grade	94 (24.80)	25 (26.32)	25 (26.60)	26 (27.08)	18 (19.15)		
2nd grade	105 (27.70)	25 (26.32)	21 (22.34)	28 (29.17)	31 (32.98)		
3rd grade	86 (22.69)	24 (25.26)	27 (28.72)	16 (16.67)	19 (20.21)		
Percent eligible for lunch program	269 (70.98)	68 (71.58)	68 (72.34)	65 (67.71)	68 (72.34)	0.87	.990
Percent with an IEP	88 (23.22)	23 (24.21)	21 (22.34)	21 (21.87)	23 (24.47)	0.27	.965
Screening characteristic SSBD							
Percent ranked 1st	257 (67.81)	59 (62.11)	67 (71.28)	69 (71.88)	62 (65.96)	2.81	.422
Critical Events Index	8.23 (3.13)	8.42 (3.32)	8.43 (2.91)	7.93 (3.45)	8.16 (2.82)	0.55	.647
CBCL externalizing behavior	25.81 (8.80)	26.60 (8.08)	25.05 (8.83)	26.61 (9.39)	24.97 (8.83)	1.03	.378

Note. Reported test statistics are *F* for continuous and χ^2 for dichotomous measures.

student compliance was excellent (*M* [*SD*] = .99 [.03]).

hB Fidelity and Dosage Data. Across 289 audio samples, average technical skills scores (*M* [*SD*] = 3.77 [0.47]), relational skill scores (*M* [*SD*] = 3.74 [0.73]), and the ratio of reflections-to-questions (*M* [*SD*] = 1.88 [1.8] 3) were above "basic competency" thresholds. The average percentage of complex reflections used during hB sessions was above "proficiency" thresholds (*M* [*SD*] = 63.53 [24.12]). Technical skills scores were above "basic competency" thresholds for 97% of the 289 sessions. Relational skills scores were above the cutoff for 80% of sessions; the use of complex reflections was above the cutoff for 86% of sessions; and the use of reflections-to-questions was above the cutoff for 59% of sessions.

On average, participants in the hB condition completed two hB steps (*SD* = 1.29). Over half (55%) completed three or four steps; an additional 6% completed two steps; and 14% only completed the first step (i.e., engage in values discovery). Nearly one-quarter of participants (24%) did not participate in any steps, meaning they did not receive any intervention support despite randomization to the hB condition.

Program Effectiveness and Participant Engagement. Coaches who delivered the FSN

Table 2. Unadjusted Baseline and Post-Intervention Means by Condition.

	Control (n = 95)		FSN only (n = 94)		HB only (n = 96)		FSN + hB Only (n = 94)	
	Baseline M (SD)	Post M (SD)	Baseline M (SD)	Post M (SD)	Baseline M (SD)	Post M (SD)	Baseline M (SD)	Post M (SD)
Pro-Social Behavior								
SSBD-ABI	28.48	30.67	29.44	35.15	29.39	31.92	29.70	34.41
	(6.38)	(8.08)	(6.91)	(9.11)	(7.36)	(8.36)	(5.04)	(8.07)
SSiS-SS-Teacher	73.31	75.92	74.24	84.43	73.31	78.33	72.95	82.97
	(9.32)	(12.26)	(9.51)	(12.92)	(11.68)	(13.22)	(9.28)	(12.00)
SSiS-SS-Parent	79.73	83.84	79.45	85.70	80.35	86.94	81.10	88.08
	(13.48)	(13.00)	(15.23)	(13.63)	(15.93)	(17.64)	(16.27)	(17.95)
Problem behavior								
SSBD-MBI	38.82	35.95	37.48	31.25	36.43	32.43	37.86	31.28
	(7.44)	(8.00)	(7.77)	(8.99)	(7.22)	(8.29)	(6.43)	(9.10)
SSiS-PB-Teacher	135.88	133.20	135.37	124.53	132.60	125.45	134.71	124.22
	(11.68)	(14.94)	(13.14)	(16.06)	(13.09)	(13.43)	(12.43)	(14.51)
SSiS-PB-Parent	131.70	126.04	128.04	120.04	128.49	123.32	128.96	125.71
	(13.75)	(15.46)	(18.64)	(17.38)	(16.58)	(17.53)	(17.23)	(19.36)
Student-teacher	37.88	36.43	39.58	35.55	37.42	34.08	37.95	32.99
conflict	(9.00)	(9.94)	(9.94)	(12.02)	(10.58)	(10.86)	(10.18)	(11.19)
Academic								
Academic	87.39	88.35	86.79	88.38	87.24	88.30	85.95	87.98
competence	(15.80)	(16.98)	(14.43)	(13.42)	(14.71)	(16.02)	(15.30)	(16.86)
Academic	56.80	60.46	57.64	66.65	60.05	65.31	55.40	68.32
engaged time	(17.83)	(17.24)	(15.85)	(17.68)	(19.69)	(19.31)	(17.91)	(17.25)

intervention reported on the FSN intervention's effectiveness and the teacher's engagement in the intervention process. Mean FSN effectiveness scores were 4.35 (SD = 0.70) and mean teacher engagement scores were 4.12 (SD = 0.67). FSN effectiveness scores and teacher engagement scores were correlated (r = .48). Coaches who delivered the hB intervention reported on the hB intervention's effectiveness and the family's engagement in the intervention process. Mean effectiveness scores were 3.80 (SD = 0.82) and mean engagement scores were 3.66 (SD = 0.83) on a 5-point Likert scale. Effectiveness and engagement scores were also highly correlated (r = .83).

Alliance and Satisfaction Scores. Parents in the FSN + hB condition (M [SD] = 4.43 [0.62]) and parents in the hB condition (M [SD] = 4.27 [0.68]) reported similar mean alliance scores. Parents in the FSN only condition reported slightly lower levels of alliance with the coach (M [SD] = 4.05 [0.68]). Teacher scores reporting alliance with the FSN coach were comparable for teachers in the FSN + hB condition (M [SD] = 4.76 [0.54]) and the FSN only condition (M [SD] = 4.70 [0.60]). Coach-reported alliance with the teacher was high across the FSN + hB (M [SD] = 4.55 [0.67]) and FSN only (M [SD] = 4.42 [0.79]) conditions. Coach-reported alliance scores with the parent were higher for the FSN + hB condition (M [SD] = 3.97 [1.02]) than the hB only (M [SD] = 3.63 [1.20]) condition.

Parents randomized to the FSN + hB condition reported higher mean levels of overall satisfaction (M [SD] = 4.31 [0.72]) as compared to parents randomized to the hB only (M [SD] = 4.06 [0.66]) and the FSN only (M [SD] = 4.10 [0.61]) conditions. Mean teacher satisfaction was high overall (M [SD] = 3.92 [0.65]). Satisfaction ratings were comparable for teachers randomized to FSN + hB (M [SD] = 3.89 [0.63]) and teachers randomized to the FSN only condition (M [SD] = 3.94 [0.68]).

Table 3. Condition Effects and Hedges' *g* Effect Sizes.

		FSN			hB			FSN + hB		
	Intercept	*b*	*SE*	*g*	*b*	*SE*	*g*	*b*	*SE*	*g*
Pro-social behavior										
SSBD-ABI	30.95	4.07***	1.15	.51	0.81	1.03	.11	3.18**	1.11	.42
SSiS-SS-Teacher	75.91	7.98***	1.54	.75	2.12	1.49	.21	7.26***	1.58	.66
SSiS-SS-Parent	83.71	2.85	1.69	.24	3.30	1.90	.25	3.48	1.94	.26
Problem behavior										
SSBD-MBI	35.36	−3.99**	1.16	.50	−2.18	1.13	.28	−4.08***	1.15	.51
SSiS-PB-Teacher	132.43	−8.35***	1.91	.64	−5.61**	1.77	.46	−8.24***	1.79	.67
SSiS-PB-Parent	124.33	−2.93	2.06	.21	−1.58	2.24	.10	1.42	2.20	.09
STRS conflict subscale	36.80	−2.04	1.30	.23	−1.93	1.21	.23	−3.34**	1.28	.38
Academic										
SSiS-AC-Teacher	87.28	0.98	1.08	.13	0.05	1.06	.01	1.66	1.15	.21
SSBD-AET	60.02	6.47*	2.55	.37	3.33	2.47	.20	9.40***	2.43	.56

*** $p < .001$, ** $p < .01$, * $p < .05$.

Posttest Differences on Outcome Measures

Table 2 summarizes unadjusted baseline and post-intervention means for each condition and Table 3 details the condition effects from the regression models and effect sizes for each intervention as compared to the control condition. In Table 3, the intercept represents the mean post-test score of the control condition on each outcome and the intervention effect (i.e., b) represents the change in the posttest score attributable to the intervention (i.e., posttest adjusted mean). As compared to the control condition, students randomized to the FSN and FSN + hB conditions had statistically significant improvements in adaptive behavior and social skills in the school setting; reductions in maladaptive behavior and problem behavior in the school setting; and improvements in academic engagement. Additionally, students randomized to the FSN + hB condition had statistically significant reductions in teacher–student conflict. Students in the hB condition had statistically significant reductions in problem behavior in the school setting. All effects remained significant after adjusting for the B-H correction. Within the home setting, effect sizes were small for the three intervention conditions. Hedges' g effect sizes for improvements in social skills and reductions in problem behavior in the home setting ranged from .24 to .26 and .09 to .21, respectively, for the three interventions. For the FSN and FSN + hB conditions, effect sizes were in the medium range across the majority of school-based outcomes and in the small range for the hB condition.

Table 4 summarizes pooled estimates (across the 40 imputed datasets) of the percent of participants who improved on each parent-reported and teacher-reported DSM-oriented diagnostic outcome from baseline to posttest, as well as the odds ratio for each intervention effect. Students randomized to FSN were 3 times and 2.3 times more likely to make statistically significant improvements on teacher-reported ADHD problems and CD problems, respectively, than students in the control condition. For the combined FSN + hB condition, students were 5 times more likely than students randomized to the control condition to improve on teacher-reported ADHD problems and 3.2 times more likely to improve on teacher-reported CD problems. Finally, students in the hB condition were 2.3 times more likely to improve on teacher-reported ADHD problems.

Practical Significance of Intervention Effects

The WWC improvement index, which represents the expected change in percentile rank

Table 4. Improvement on Diagnostic Outcomes by Condition.

	Control	FSN		hB		FSN + hB	
	Improved *n* (%)	Improved *n* (%)	*OR*(CI)	Improved *n* (%)	*OR*(CI)	Improved *n* (%)	*OR*(CI)
ADHD							
Teacher report	15 (15.79)	34 (36.17)	3.01** (1.44, 6.27)	29 (30.21)	2.28* (1.08, 4.82)	45 (47.87)	4.96*** (2.39, 10.30)
Parent report	31 (32.63)	38 (40.43)	1.37 (0.72, 2.60)	38 (39.58)	1.36 (0.70, 2.61)	33 (35.11)	1.13 (0.58, 2.20)
ODD							
Teacher report	26 (27.37)	29 (30.85)	1.15 (0.60, 2.23)	25 (26.04)	0.91 (0.46, 1.78)	39 (41.49)	1.83 (0.97, 3.47)
Parent report	48 (50.53)	54 (57.45)	1.34 (0.73, 2.45)	45 (46.88)	0.88 (0.48, 1.59)	48 (51.06)	1.01 (0.56, 1.82)
CD							
Teacher report	18 (18.95)	34 (36.17)	2.33* (1.14, 4.77)	27 (28.13)	1.60 (0.77, 3.31)	41 (43.62)	3.18** (1.56, 6.47)
Parent report	42 (44.21)	46 (48.94)	1.18 (0.65, 2.16)	37 (38.54)	0.77 (0.41, 1.48)	39 (41.49)	0.91 (0.48, 1.69)

OR(CI) = Odds ratio and 95% confidence interval.

for an average student in the control group if that child had received the intervention, showed improvement for each of the three intervention conditions. For the FSN condition, the mean improvement index score was + 18.89 percentile points (range = + 9.64 to + 27.43), + 14.99 percentile points (range = + 8.17 to + 23.74), and + 9.80 percentile points (range = + 5.22 to + 14.37) for prosocial behavior, problem behavior, and academic outcomes, respectively. For the hB condition, the mean improvement index score was + 7.53 percentile points (range = + 4.53 to + 9.93) for improvements in student prosocial behavior; + 10.49 percentile points (range = + 4.07 to + 17.67) for reductions in student problem behavior; and + 4.01 percentile points (range = + 0.28 to + 7.74) for improvements in academic outcomes. Finally, for the combined FSN + hB condition, the mean improvement score for prosocial behavior was + 16.99 (range = + 10.24 to + 24.65); the mean improvement score for reductions in problem behavior was + 13.86 (range = + 3.72 to + 24.81); and the mean improvement score for academic outcomes was + 14.73 (range = + 8.24 to + 21.23). Across all intervention conditions, improvements were greater for school-based rather than home-based outcomes. With respect to outcomes in the home setting, mean improvement scores were + 8.91 percentile points for the FSN condition, + 7.00 percentile points for the hB condition, and + 6.98 percentile points for the FSN + hB condition. For school-based outcomes, mean improvement scores were + 8.37 percentile points for the hB condition, + 16.91 percentile points for the FSN condition, and + 18.48 percentile points for the FSN + hB condition. With respect to school-based outcomes, the combined FSN + hB condition, in particular, had higher improvement index scores for reductions in teacher-student conflict (+ 14.77 vs. + 9.00) and improvements in academic engagement (+ 21.23 vs. + 14.37) as compared to the FSN only condition.

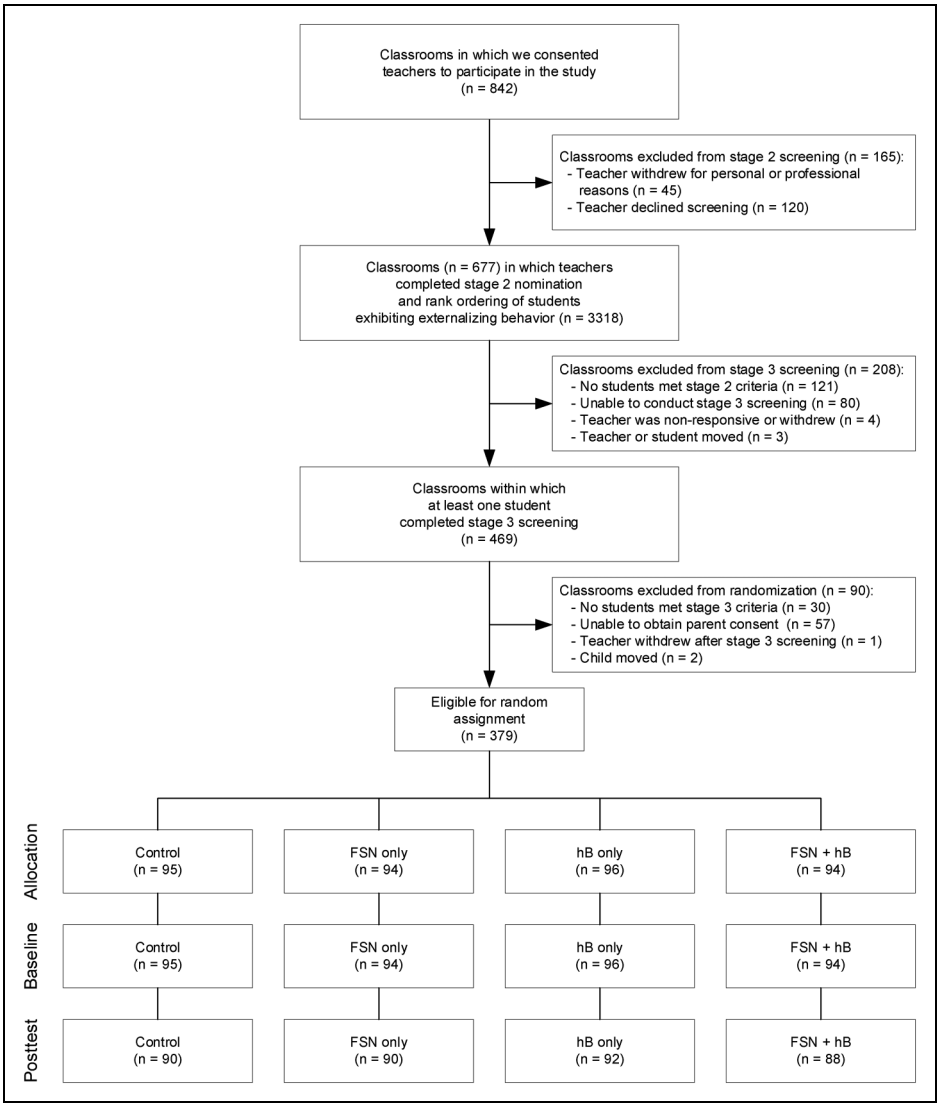

Figure 1. CONSORT Diagram

Discussion

This study addressed disruptive behavior in childhood, which is increasingly pervasive (U.S. Department of Education, 2018), associated with comorbid externalizing behavior problems (Bradshaw et al., 2010), and often results in longterm negative consequences without early intervention (Zaim & Harrison, 2020). The study illustrates the potential benefit of interventions that involve the teacher as the primary program implementer for students requiring intensive support (Kaminski & Claussen, 2017), as well as parental interventions that directly involve children (Dretzke et al., 2009; Granski et al., 2020). Finally, the study contributes to our knowledge of the hB intervention. Although there is some prior evidence that hB is efficacious when implemented with First Step to Success (See Frey et al., 2015), the effect of the intervention had not previously been established within the context of an RCT.

Our analytic approach provided a nuanced picture of our outcomes. Consistent with previous First Step studies, there was compelling

data demonstrating the efficacy of FSN (See Feil et al., 2014; Feil et al., 2021; Sumi et al., 2013; Walker et al., 1998; Walker et al., 2009). Specifically, whether examining participants in the FSN only intervention and the control group, statistically significant differences and small to medium effect sizes were observed across multiple indicators for prosocial and problem behavior, as well as AET, in the school setting. The categorical analysis of DSM-oriented diagnostic outcomes also indicated that those in the FSN condition were 3 times more likely to make statistically significant improvement on ADHD problems and over 2 times more likely on CD problems. Finally, for the FSN condition, the improvement indices demonstrated impressive percentile improvements across prosocial behavior, problem behavior, and academic outcomes. The significant findings on AET were particularly important because the outcome was measured via direct observation and is related to academic performance.

The results for hB were modest, with the only statistically significant reductions in problem behavior for students in the hB only condition occurring in the school setting, rather than the home setting where we anticipated them to be most robust. Yet, the diagnostic analysis provided a more optimistic picture of the potential efficacy of hB with higher percentages of children from the hB only condition moving from at-risk status with regard to ADHD symptoms at baseline to the normative range at posttest. Although these results appear inconsistent with previous RCTs demonstrating parent-only interventions yielding mean effect sizes over 1.0, it is important to note that the hB component requires relatively little time or resources to implement compared to many other parent interventions (Epstein et al., 2015). Further, in these studies the parent intervention is implemented by mental health clinicians, whereas our implementers were school personnel. It is also worth noting, however, that nearly 25% of the parents assigned to a hB condition did not actually begin the intervention.

Another important finding is related to the combined intervention condition, which demonstrates that adding a brief, parent-focused intervention to FSN, was beneficial. Specifically, only those in the FSN + hB condition had statistically significant reductions in teacher–student conflict, and several outcomes demonstrated more robust gains for students in the FSN + hB condition than the FSN only or hB only conditions, irrespective of the analytic approach. The value of adding an easy to implement home intervention to an established school intervention is noteworthy. Because hB was designed to be deployed by school-based personnel, examining its efficacy in combination with other teacher, child, or teacher–child component interventions may also be warranted (Vidair et al., 2014). Examining the efficacy of the intervention with a less severe (i.e., targeted vs. indicated level) sample, alone and in combination with other interventions, could also be useful, as would examining maintenance gains over long-term follow-up.

Finally, all of our reported process data were favorable for both FSN and hB. Despite its long history of demonstrating favorable results in these areas, this is the first study with the revised FSN program variation conducted in the elementary grades, suggesting that the intervention's revision changes did not negatively influence implementation or social validity. For hB, these results were encouraging, and this is the first time they have been studied in the context of a large RCT. Quality of hB implementation—indicated by MI quality—is also noteworthy.

There were some limitations to the current study that are important to recognize. First, although the training we provided coaches in both conditions was robust, and the fidelity of implementation was a strength of the study, we did not establish fidelity criteria that needed to be met prior to being assigned to students, teachers, or parents to work with. Inclusion of such criteria would have strengthened the study and should be utilized in future research. Second, Little's MCAR test to assess missing data has low power and is susceptible to Type II errors (Enders, 2010). Third, we included the conflict subscale of the STRS but not the closeness subscale. Future research should include the

closeness subscale, as well as an analytic strategy to determine if it has a moderating effect on the degree of conflict. Finally, sustained effect is the goal of all intervention research. In the current study, we did collect 6-month follow-up data from parents (only). We will present the results of that data in a subsequent manuscript involving a contemporary mediation analysis involving those in the hB only and hB plus FSN conditions. Additionally, documenting sustained effects of all of the outcomes should be a focus of future research.

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. University of Vermont, Research Center for Children, Youth, & Families.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Blanchard, L. T., Gurka, M. J., & Blackman, J. A. (2006). Emotional, developmental, and behavioral health of American children and their families: A report from the 2003 National Survey of Children's Health. *Pediatrics*, 117(6), e1202–e1212. https://doi.org/10.1542/peds.2005-2606

Bradshaw, C. P., Schaeffer, C. M., Petras, H., & Ialongo, N. (2010). Predicting negative life outcomes from early aggressive–disruptive behavior trajectories: Gender differences in maladaptation across life domains. *Journal of Youth & Adolescence*, 39(8), 953–966. https://doi10.1007/s10964-009-9442-8

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, 86(2), 127–137

Dretzke, J., Davenport, C., Frew, E., Barlow, J., Stewart-Brown, S., Bayliss, S., Taylor, S. S., Sandercock, J., & Hyde, C. (2009). The clinical effectiveness of different parenting programmes for children with conduct problems: A systematic review of randomised controlled trials. *Child & Adolescent Psychiatry & Mental Health*, *3*, 1–10. https://doi.org/10.1186/1753-2000-3-7

Enders, C. (2010). *Applied missing data analysis*. The Guilford Press.

Epstein, R. A., Fonnesbeck, C., Potter, S., Rizzone, K. H., & McPheeters, M. (2015). Psychosocial interventions for child disruptive behaviors: A meta-analysis. *Pediatrics*, 136(5), 947–960. https://doi.org/10.1542/peds.2015-2577

Feil, E., Frey, A. J., Walker, H. M., Small, J., Seeley, J., Golly, A., & Forness, S. R. (2014). The efficacy of a home-school intervention for preschoolers with challenging behaviors: A randomized controlled trial of preschool first step to success. *Journal of Early Intervention*, 36, 151–170. https://doi.org/10.1177/1053815114566090

Feil, E. G., Walker, H. M., Frey, A. J., Seeley, J., Small, J. W., Golly, A., Lee, J., & Forness, S. R. (2021). Efficacy validation of the revised first step program: A randomized controlled trial. *Exceptional Children*, 87(2), 183–198. https://doi.org/10.1177/0014402920924848

Forness, S. R., Freeman, S. F., Paparella, T., Kauffman, J. M., & Walker, H. M. (2012). Special education implications of point and cumulative prevalence for children with emotional or behavioral disorders. *Journal of Emotional and Behavioral Disorders*, 20(1), 4–18. https://doi.org/10.1177/1063426611401624

Forness, S. R., Walker, H. M., & Serna, L. A. (2014). Establishing an evidence-base: Lessons learned from implementing randomized controlled trials for behavioral and pharmacological interventions. In H. M. Walker, & F. M. Gresham (Eds.), *Handbook of evidence-based practices for emotional and behavioral disorders* (pp. 567–582). The Guilford Press.

Frey, A. J., Lee, J., Small, J. W., Walker, H. M., & Seeley, J. R. (2017). Motivational interviewing training and support for school-based applications. *Emotional & Behavioral Disorders in Youth*, *17*(4), 86–92. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6345406/pdf/nihms982106.pdf

Frey, A. J., Small, J. W., Lee, J., Crosby, S., Seeley, J. R., Forness, S., & Walker, H. M. (2019). Homebase: Participation, engagement, alliance, and social validity of a motivational parenting intervention. *Children & Schools*, Advance online publication. http://doi.org/10.1093/cs/cdz016

Frey, A. J., Small, J. W., Lee, J., Walker, H. M., Seeley, J. R., Feil, E. G., & Golly, A. (2015). Expanding the range of the first step to success intervention: Tertiary-level support for children, teachers, and families. *Early Childhood Research Quarterly*, *30*, 1–11. https://doi.org/10.1016/j.ecresq.2014.05.002

Granski, M., Javdani, S., Anderson, V. R., & Caires, R. (2020). A meta-analysis of program characteristics for youth with disruptive behavior problems: The moderating role of program format and youth gender. *American Journal of Community Psychology*, *65*(1–2), 201–222. https://doi.org/10.1002/ajcp.12377

Gresham, F. M., & Elliott, S. N. (2008). *Social skills improvement system rating scales.* NCS Pearson.

Group, M. A. C. S. R. (2002). A cognitive-ecological approach to preventing aggression in urban settings: Initial outcomes for high-risk children. *Journal of Consulting and Clinical Psychology*, *70*, 179–194. https://doi.org/10.1037/0022-006X.70.1.179

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*, 23–34. https://doi.org/10.20982/tqmp.08.1.p023

Kaminski, J. W., & Claussen, A. H. (2017). Evidence base update for psychosocial treatments for disruptive behaviors in children. *Journal of Clinical Child & Adolescent Psychology*, *46*(4), 477–499. https://doi.org/10.1080/15374416.2017.1310044

Miller, W. R., & Rollnick, S. (2012). *Motivational interviewing: Helping people change*. The Guilford Press.

Moyers, T. B., Rowell, L. N., Manuel, J. K., Ernst, D., & Houck, J. M. (2016). The motivational interviewing treatment integrity code (MITI 4): Rationale, preliminary reliability and validity. *Journal of Substance Abuse Treatment*, *65*, 36–42. https://doi.org/10.1016/j.jsat.2016.01.001

Muthèn, L. K., & Muthèn, B. O. (1998−2010). *Mplus user's guide*, (6th ed.). Psychological Assessment Resources.

Noltemeyer, A. L., Marie, R., McLoughlin, C., & Vanderwood, M. (2015). Relationship between school suspension and student outcomes: A meta-analysis. *School Psychology Review*, *44*(2), 224–240. https://doi.org/10.17105/spr-14-0008.1

Pianta, R. (2001). *Student-teacher relationship scale. Professional manual. Lutz.* Psychological Assessment Resources.

Reinke, W. M., Eddy, J. M., Dishion, T. J., & Reid, J. B. (2012). Joint trajectories of symptoms of disruptive behavior problems and depressive symptoms during early adolescence and adjustment problems during emerging adulthood. *Journal of Abnormal Child Psychology*, *40*(7), 1123–1136. https://doi.org/10.1007/s10802-012-9630-y

Seeley, J. R., Severson, H. H., & Fixsen, A. A. M. (2014). Empirically based targeted prevention approaches for addressing externalizing and internalizing behavior disorders within school contexts. In H. M. Walker, & F. M. Gresham (Eds.), *Handbook of evidence-based practices for emotional and behavioral disorders* (pp. 307–323). The Guilford Press.

Small, J. W., Frey, A., Lee, J., Seeley, J. R., Scott, T. M., & Sibley, M. H. (2020). Fidelity of motivational interviewing in school-based intervention and research. *Prevention Science*, 1–10. https://doi.org/10.1007/s11121-020-01167-7

Sprague, J., & Golly, A. (2013). *BEST Behavior. Building positive behavior support in schools* (2nd ed.). Sopris Learning.

Sumi, W. C., Woodbridge, M. W., Javitz, H. S., Wagner, M., Rouspil, K., & Severson, H. (2013). Assessing the effectiveness of first step to success: Are short-term results the first step to long-term behavioral improvements? *Journal of Emotional and Behavioral Disorders*, *21*, 66–78. https://doi.org/10.1177/1063426611429571

U.S. Department of Education, Institute for Education Sciences, National Center for Education Evaluation and Regional Assistance. *What Works.* Clearinghouse.

U.S. Department of Education, Office of Civil Rights (2018). 2015-2016 civil rights data collection: School climate and safety. https://www2.ed.gov/about/offices/list/ocr/docs/school-climate-and-safety.pdf

Vidair, H. B., Sauro, D., Blocher, J. B., Scudellari, L. A., & Hoagwood, K. E. (2014). Empirically supported school-based mental health programs targeting academic and mental health functioning: An update. In H. M. Walker, & M. G. Gresham (Eds.), *Handbook of evidence-based practices for emotional and behavioral disorders* (pp. 15–53). The Guilford Press.

Walker, H., Feil, E. G., Frey, A., Small, J., Seeley, J., Golly, A., Crosby, S., Lee, J., Forness, S., Sprick, M., Coughlin, C., & Stiller, B. (2018). First Step Next: An updated version of the First Step to success early intervention program. *Grantee Submission*, 3(1), 89–109. https://files.eric.ed.gov/fulltext/ED602357.pdf

Walker, H., Kavanagh, K., Stiller, B., Golly, A., Feil, E., & Severson, H. (1997). *First step to success: Helping young children overcome antisocial behavior*. Sopris West, Inc.

Walker, H. M., Forness, S. R., & Lane, K. L. (2014). Design and management of scientific research in applied school settings.

In B. G. Cook, M. Tankersley, & L. J. Timothy's (Eds.), *Advances in learning and behavioral disabilities*, *27*, (141–169). Emerald Publishing.

Walker, H. M., Kavanagh, K., Stiller, B., Golly, A., Severson, H. H., & Feil, E. G. (1998). First step to success: An early intervention approach for preventing school antisocial behavior. *Journal of Emotional & Behavioral Disorders*, *6*(2), 66–81. https://doi.org/10.1177/106342669800600201

Walker, H. M., Seeley, J. R., Small, J., Severson, H. H., Graham, B. A., Feil, E. G., & Forness, S. R. (2009). A randomized controlled trial of the first step to success early intervention: Demonstration of program efficacy outcomes in a diverse, urban school district. *Journal of Emotional and Behavioral Disorders*, *17*(4), 197–212. https://doi.org/10.1177/1063426609341645

Walker, H. M., Severson, H. H., & Feil, E. G. (2014). Systematic screening for behavior disorders (SSBD): Administrator's Guide. *Universal Screening for preK*, 9. Ancora Publishing.

Walker, H., Stiller, B., Coughlin, C., Golly, A., Sprick, M., & Feil, E. (2015). *First step next: Coaches guide*. Pacific Northwest Publishing, Inc.

Webster-Stratton, C., Reid, M. J., & Hammond, M. (2004). Treating children with early-onset conduct problems: Intervention outcomes for parent, child, and teacher training. *Journal of Clinical Child and Adolescent Psychiatry*, *33*(1), 105–124. https://doi.org/10.1177/1063426609341645

Zaim, N., & Harrison, J. (2020). Pre-school mental health disorders: A review. *International Review of Psychiatry*, *32*(3), 189–201. https://doi.org/10.1080/09540261.2019.1692793

Acknowledgements

This article was made possible, in part, by the support of the Jefferson County (Louisville, KY) Public Schools; Fayette County (Lexington, KY) Public Schools, Jessamine County (Nicholasville, KY); and Spencer County (Taylorsville, KY) Public Schools. Opinions contained in this report/article reflect those of the authors and do not necessarily reflect those of the participating districts.

Declaration of Conflicting Interests

Hill Walker and Edward Feil are two of the authors of the First Step Next intervention. The authors declared no other potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Institute of Education Sciences (grant number R324A150179).

ORCID iDs

Andy J. Frey https://orcid.org/0000-0002-7493-5728
Edward G. Feil https://orcid.org/0000-0002-1004-8900
Dana Cohen Lissman https://orcid.org/0000-0003-0934-0813

Open Science Practices

For publishing data and materials, this article earned badges for Open Data and Open Materials. The public content for the data, code, and materials may be retrieved from https://nam12.safelinks.protection.outlook.com/?url=http%3A%2F%2Fwww.openicpsr.org%2Fopenicpsr%2Fproject%2F145801%2Fversion%2FV1%2Fview&data=04%7C01%7Cec_sagepub%40novatechset.com%7Cc8038226c93149ef797f08d9c66b8846%7Ca03a7f6cfbc84b5fb16bf634dbe1a862%7C1%7C0%7C637758987428464261%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000&sdata=3kgcvDbFYL9bJPPqpY%2B1quXHql%2FOwBOm0E2LWOqoZw8%3D&reserved=0.